WAIS Inc

Wide Area Information Servers
690 Fifth Street
San Francisco California 94107
415-356-5400
FAX 415-356-5444

FAX Cover Sheet

Date: MAY 11, 1995
Attn: JUDITH BARKER

Phone: 386-1605
FAX: 386-3531

From: Dia Cheney/Marketing

Phone: 415-356-5413
FAX: 415-356-5444

Pages Transmitted (including cover sheet): 8

COMMENTS:

Hi Judith,

Here are the 7 diagrams for our technical documentation.

Please call me if you have any questions - Thanks!

Thanks,

Dia

Dia Cheney
dia@wais.com
415-356-5413

Tech. Description p1.

fix corners and joins

1

Introduction

What is WAIS?

WAIS™ (Wide Area Information Servers™) is a network publishing system designed to help users find information over a computer network by simply asking questions. The questions may be expressed in natural language or use literal phrases, Boolean syntax, or specify field values. The information sources may be local or remote. WAIS software allows users to search for and retrieve documents from information sources all over the world.

As organizations become flatter and more geographically dispersed, the WAIS network publishing system offers an efficient method for accessing information electronically over interconnected local and wide-area networks, thereby greatly reducing printing and distribution time and expenses.

The WAIS Architecture

The WAIS software architecture has four main components: the client, the server, the database, and the protocol, as shown in Figure 1. The WAIS client is a user-interface program that sends search and retrieval requests to local or remote servers. Clients are available for most popular desktop environments. The WAIS server is a program that services client requests. Servers are available on a variety of UNIX platforms. The server generally runs on a machine containing one or more information sources, or WAIS databases. The WAIS protocol is used to connect WAIS clients and servers and is based on the NISO Z39.50 Information Retrieval Service and Protocol Standard.

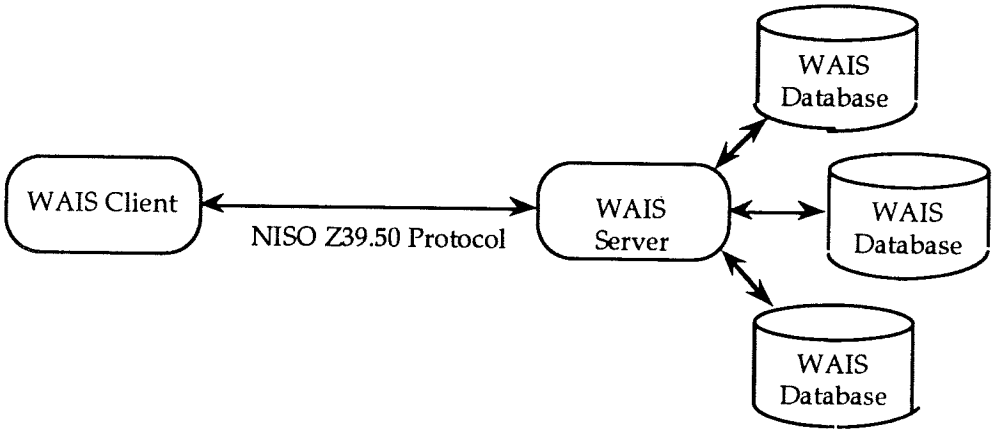

Figure 1: The WAIS Architecture

pointers into the document table corresponding to the documents that contain that word. The information from the inverted file is used to look in the document table, which gives a pointer to the headline table, which in turn gives a pointer to the filename table. Finally, the information from the filename table is used to access the original data. A list of headlines and relevance scores is returned to the client process for display to the end user.

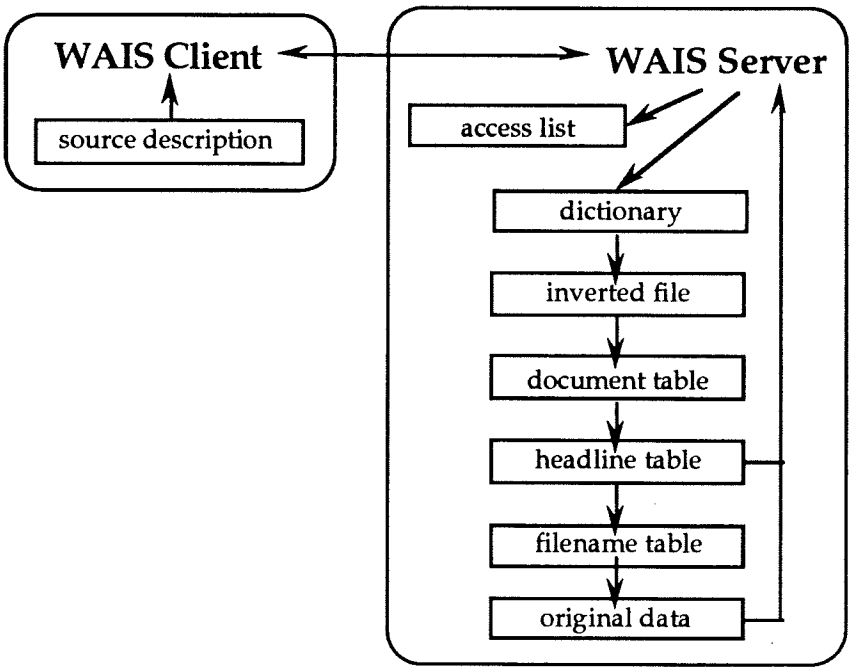

Figure 3: How the WAIS index is used during a search.

Incremental Indexing

The WAIS indexer offers incremental indexing. Incremental indexing allows you to add, modify, or delete WAIS database documents without reindexing the entire database and without suspending user service. Incremental indexing modifies the WAIS index to reflect data changes since the last time the data was indexed. This capability is especially important for network publishers whose data changes often, and whose database size is large.

Customizable Stopwords

A stopword is a frequently used word that, when encountered in a user question, is ignored. For example, since the word "the" commonly appears throughout the English language, it does not help distinguish between documents. Thus it is typically regarded as a stopword. The WAIS indexer includes a list of approximately 300 standard stopwords which can be specially customized for each WAIS database.

Stemming

Stemming is a technique used to automatically derive variations of a queried word. These variations are then used as part of the search. If stemming is used, then when a data set is indexed, word stems are indexed where possible. For example, "dancing," "danced," and "dancer"

The goal of the WAIS network publishing system is to create an open architecture of information servers and clients by using a standard computer-to-computer protocol that enables clients to communicate with servers.

The WAIS client-server architecture has many advantages:

Scalability

Its distributed nature allows anyone to set up their own server and become a network publisher. The system can handle thousands of information sources on internets that span the globe, all searchable using standard software.

Efficiency

Current personal computers are high-powered and responsive to the user, and server machines have increased storage capacity and the ability to simultaneously service many users. The client-server architecture lets the client machine interact with the user as a native application on its platform. For example, a WAIS client for Microsoft Windows is a true Windows application and behaves as Windows users expect. Contrast this with most on-line services where a remote server controls what the user sees. The WAISserver, on the other hand, receives its questions in a standard format from all clients and can handle requests without having to recode the response for individual client programs.

Global Communication

The distances involved in global client-server applications often equate to a minimum delay of about one second. Dialup, low-speed lease lines, and wireless connections are typically the most cost-effective means users have to connect to wide-area networks. If information is transmitted on a character-by-character basis over a slow link, the delay between each keystroke could be intolerable. A client-server system can hide much of this delay by packaging up a significant parcel before sending it from the client to the server.

What is a WAIS Network Publisher?

A WAIS network publisher is an information provider that supplies both a WAIS database and a WAIS server, as shown in Figure 2.

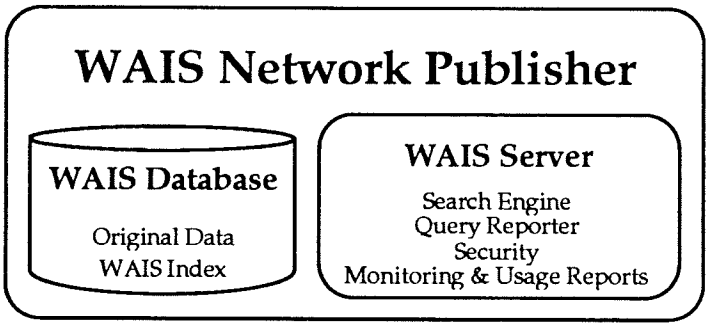

Figure 2: The WAIS Network Publisher

A WAIS database is made up of the publisher's original data collection and a WAIS index to facilitate fast search and retrieval of this data. The WAISserver system is composed of a search engine, a query reporter, a security system, and a monitoring and usage reporting facility. The

Tech Description

6

The WAIS Forwarder

The WAIS Forwarder product, in conjunction with a "firewall" machine, provides access to external WAIS servers from within secure environments. The WAIS Forwarder is appropriate for secure sites connected to an external network, such as the Internet, through a firewall machine.

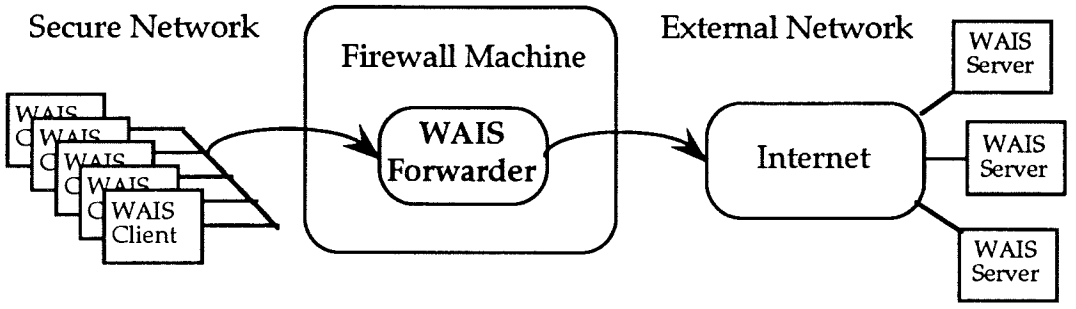

Figure 4: Configuration of the WAIS Forwarder

As shown in Figure 4, a firewall is a machine that connects a secure network to an external network. Information from one network destined for the other network must pass through the firewall machine. A forwarder is a software program running on the firewall machine that permits two application programs executing on either side of the firewall to communicate with each other. The forwarder allows machines on the secure network to access the services available on machines in the external network.

In a client-server application such as WAIS, a client contacts the forwarder on the firewall machine and the forwarder contacts the outside servers. Secure machines can open connections to Internet servers transparently by sending the request to the forwarder which automatically passes the request onto the external service. External machines cannot open connections to the forwarder, thus forming a one-way security system.

The WAIS Forwarder provides a secure network with all the benefits of the Internet WAIS servers without opening the secure network to external traffic. All WAIS functions are supported through the forwarder including the Directory of Servers, searching, and retrieval of text, images, and other formats. Because the WAIS Forwarder also forwards the IP address of the requesting client machine, databases using WAIS Inc. servers will continue to provide access-list security. In addition, the WAIS Forwarder optionally logs transaction statistics, enabling the firewall maintainer to monitor usage patterns.

The WAIS Forwarder is a software-only product that runs on many popular UNIX platforms and is easily configured and administered. In addition, the WAIS Forwarder works with all existing client software. For those that have special needs or security considerations, the product is available in source code as well as in executable form. The WAIS Forwarder can be purchased separately, or bundled with the WAIS Inc. server products. As new versions of the WAIS protocol suite come into widespread use, the package will be upgraded according to the maintenance and support agreement selected.

right, the flow of information can be traced from its original repository, to its destination in response to a client's search. Taken together, the original data and the WAIS index make up a complete WAIS database. The **waislookup** command can be used for testing the database. The server executes the **waisserver** program which references the WAIS database, and returns answers to the user's question. The server also writes logs that are summarized by the **waisreporter** program.

Now consider the alternative path, from the client, asking a question of the server. The dotted arrows illustrate how a user's question and relevant documents are fed to the server. The server examines the index, which refers it to the original data. This produces a list of headlines which the server returns to the client. The same path is followed when the client makes a retrieval request. This time however, actual records from the original data are returned.

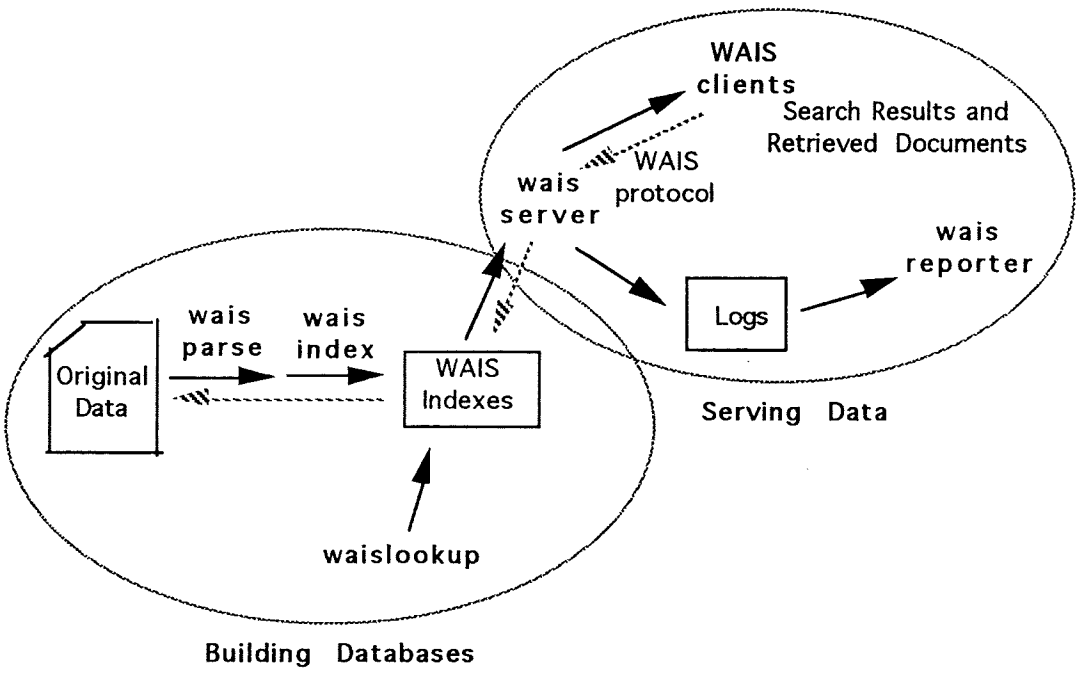

Figure 1: Components of the WAIS System

The important relationships are as follows:

- The **waisparse** and **waisindex** programs build a WAIS database from one or more data files which you provide.
- WAIS clients and servers communicate using the WAIS protocol.
- The **waisserver** program examines the WAIS index and the original data when it responds to a user's question.

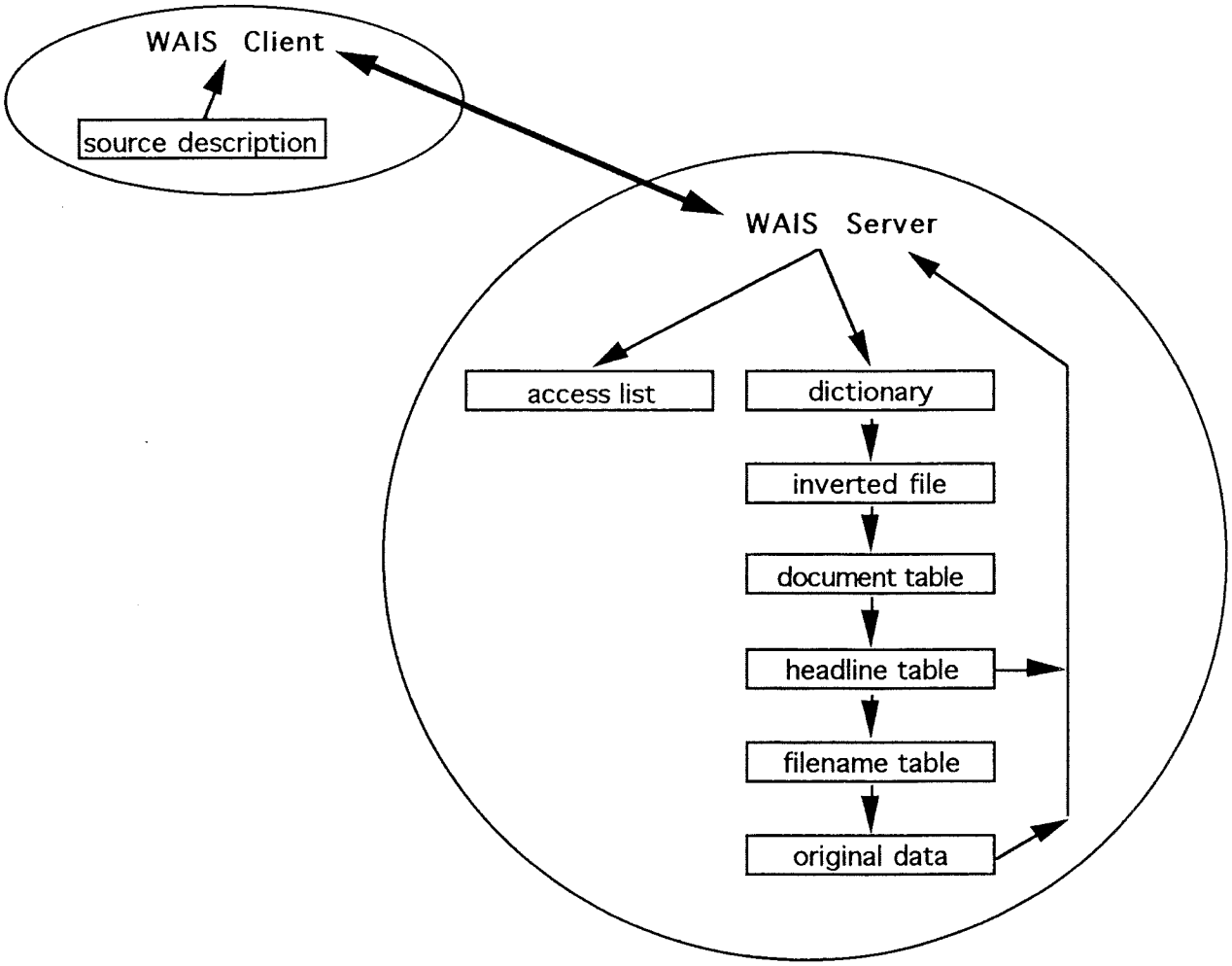

Figure 3: How index files are used during a search.

The interaction between the WAIS index files is illustrated in Figure 3. A client process uses information from the source description file to find and contact the server. The server checks the access list to make sure the client has permission to access this database. If so, the server process takes the words from the client's query and looks them up in the database's dictionary file. The dictionary file provides pointers into the inverted file, where, for each word, there is a list of pointers into the document table corresponding to the documents that contain that word. The information from the inverted file is used to look in the document table, which gives a pointer to the headline table, which in turn gives a pointer to the filename table. Finally, the information from the filename table is used to access the original data. A list of headlines and relevance scores is returned to the client process for display to the end user.